Contents lists available at Evidance Health Sciences

Evidance Health Sciences

journal homepage: www.ehealthsciences.com

eISSN: 3122-3354
Evidance Scientific Organization

Issue: 1
Volume: 1
Year: 2026

Diagnostic Accuracy of Point-of-Care and Handheld Ultrasound for Abdominal Aortic Aneurysm Detection: A Systematic Review and Meta-Analysis

Yazan J. Alalwani[a], Lamees A. Alzahrani[b], Maysam H. Alyami[b], Abdulrahman A. Alanazi[c], Mawaddah A. Alsobay[d], Ibtihaj A. Alsalhi[e], Alhawraa H. Alawamia[d], Jumanah S. Al-Siari[f], Reema A. Alharbi[g], Dalya M. Motabagani[h], Ahmed Y. Azzam[i]*

[a] College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[b] College of Medicine, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia
[c] College of Medicine, Shaqra University, Riyadh, Saudi Arabia
[d] College of Medicine, King Saud University, Riyadh, Saudi Arabia
[e] College of Medicine, King Khalid University, Abha, Saudi Arabia
[f] College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[g] College of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia
[h] College of Medicine, Vision Colleges, Riyadh, Saudi Arabia
[i] Division of Global Health and Public Health, School of Nursing, Midwifery and Public Health, University of Suffolk, Ipswich, United Kingdom

* Corresponding author. E-mail address: ahmedyazzam@gmail.com (Ahmed Y. Azzam).

ARTICLE INFO

Article type:
Systematic Review And Meta-Analysis

Access:
Open Access

Submission type:
Regular Submission

DOI:
10.65416/ehealthsci.2026.673710

Journal:
Evidance Health Sciences

Editor-in-Chief:
Eman Mohammed Alshammari, MD

eISSN:
3122-3354

King Fahad National Library Journal Registration Number:
1447/13452

Publisher:
Evidance Scientific Organization

Scientific Publisher License Registration Number in General Authority of Media Regulation, Kingdom of Saudi Arabia:
165544

Keywords:
Abdominal Aortic Aneurysm
Point-of-Care Ultrasound
Handheld Ultrasound
Diagnostic Accuracy
Emergency Department

ABSTRACT

Introduction: Abdominal aortic aneurysm (AAA) represents a life-threatening vascular condition requiring rapid diagnosis in emergency settings. Point-of-care ultrasound (POCUS) and handheld ultrasound devices (HUD) have been proposed for rapid AAA detection; however, their diagnostic accuracy remains incompletely characterized through structured meta-analysis. This systematic review and meta-analysis evaluated the diagnostic accuracy of POCUS and HUD for AAA detection.

Methods: Following PRISMA 2020 guidelines, we searched databases up to 2 November, 2025, for studies evaluating ultrasound-based AAA detection (≥3.0 cm threshold) compared to reference standards. Random-effects meta-analysis was performed for sensitivity and specificity. Hierarchical summary receiver operating characteristic (HSROC) curves, measurement agreement analysis, and publication bias assessment were conducted.

Results: Ten studies enrolling 1,111 patients were included. Emergency department POCUS demonstrated pooled sensitivity of 95.0% (95% confidence interval [CI]: 87.5-98.1%) and specificity of 97.7% (95% CI: 94.9-99.0%) across four studies (533 patients), with minimal heterogeneity ($I^2 = 0$-11%). Rupture detection sensitivity was 100% (95% CI: 84.5-100%, 21 cases). HUD showed sensitivity of 90.6% (95% CI: 67.7-97.8%) and specificity of 96.1% (95% CI: 89.9-98.6%) in one vascular clinic study (104 patients). Measurement agreement analysis revealed significant heterogeneity for POCUS ($I^2 = 98\%$) driven by inner-to-inner versus outer-to-outer measurement conventions, while HUD demonstrated excellent agreement (mean difference 0.51 mm, intraclass correlation coefficient 0.87). Publication bias assessment showed no significant asymmetry (Deeks' test P-value = 0.668).

Conclusions: POCUS have demonstrated excellent diagnostic accuracy for AAA detection with minimal heterogeneity, supporting implementation in emergency settings. HUD showed promising accuracy; however, evidence remains limited to single-study data requiring additional validation studies.

Categories: Emergency Medicine, Internal Medicine, Radiology & Imaging

* Corresponding author. E-mail address: ahmedyazzam@gmail.com (Ahmed Y. Azzam).

Received May 12, 2026, Accepted May 28, 2026, Available Online May 31, 2026

© 2026 The Author(s). Published by Evidance Scientific Organization. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. INTRODUCTION

Abdominal aortic aneurysm (AAA) remains a life-threatening vascular condition affecting around 5-10% of men aged over 65 year-old around the world. Rupture of AAA carries mortality rates exceeding 80%, with survival heavily dependent on rapid diagnosis and surgical intervention. In emergency department (ED) settings, timely AAA identification in patients presenting with abdominal pain, back pain, or syncope represents an important management priority, as delayed diagnosis significantly worsens outcomes. Standard imaging modalities including computed tomography (CT) provide definitive diagnosis; however, CT availability, radiation exposure, contrast requirements, and time delays limit utility in hemodynamically unstable patients or resource-limited settings [1-4].

Point-of-care ultrasound (POCUS) performed by emergency physicians at the bedside has been proposed as a rapid alternative for AAA detection. POCUS offers multiple advantages including immediate availability, portability, absence of radiation exposure, feasibility in unstable patients, and ability to guide urgent management decisions without delay. Early studies have suggested high diagnostic accuracy for POCUS in detecting AAA; however, previous evidence syntheses have been limited by small sample sizes, failure to account for measurement convention heterogeneity, absence of rupture detection analysis, and incomplete publication bias assessment. In addition to that, recent technological advancements have introduced handheld ultrasound devices (HUD) representing ultra-portable, low-cost alternatives to cart-based systems, with possible applications in primary care screening and resource-limited environments [5-9].

Despite growing adoption of POCUS and emerging interest in HUD for AAA evaluation, no structured meta-analysis has quantified diagnostic accuracy while addressing methodological challenges including measurement convention discordance (inner-to-inner versus outer-to-outer), operator experience heterogeneity, and rupture detection performance. Understanding the magnitude and consistency of diagnostic accuracy across different ultrasound modalities, clinical settings, and operator experience levels has important implications for guideline development, training requirements, and resource allocation decisions [2,6].

Therefore, we aimed to conduct a systematic review and meta-analysis to evaluate the diagnostic accuracy of POCUS and HUD for AAA detection compared to reference standards. Our objectives were to quantify pooled sensitivity and specificity using bivariate random-effects meta-analysis, investigate measurement agreement and sources of heterogeneity, evaluate rupture detection performance, assess publication bias through funnel plot asymmetry testing, and determine sample size adequacy through post-hoc power analysis. We hypothesized that both POCUS and HUD would have demonstrated high diagnostic accuracy with minimal heterogeneity across studies.

2. MATERIALS AND METHODS

Search Strategy and Study Selection:

This systematic review and meta-analysis was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [10]. We performed a detailed literature search in PubMed, Scopus, Web of Science, and the Cochrane Central Register of Controlled Trials from database inception up to 2 November, 2025, with focusing on English-language studies. The search strategy utilized a combination of Medical Subject Headings (MeSH) terms and free-text keywords designed to capture all relevant studies investigating ultrasound-based AAA detection. The structured search string included the following terms: ("abdominal aortic aneurysm" OR "AAA" OR "aortic aneurysm" OR "abdominal aorta" OR "infrarenal aorta" OR "aortic dilation" OR "aortic ectasia") AND ("ultrasound" OR "ultrasonography" OR "sonography" OR "point-of-care ultrasound" OR "POCUS" OR "bedside ultrasound" OR "emergency ultrasound" OR "handheld ultrasound" OR "portable ultrasound" OR "pocket ultrasound" OR "Vscan" OR "Butterfly" OR "SonoSite" OR "Lumify" OR "hand-carried ultrasound") AND ("diagnostic accuracy" OR "sensitivity" OR "specificity" OR "detection" OR "diagnosis" OR "screening" OR "identification" OR "measurement" OR "agreement" OR "validation" OR "performance"). Reference lists of all included studies and relevant reviews were manually screened to identify additional eligible studies not captured by electronic searches.

We first conducted title and abstract screening according to our eligibility criteria. Studies were included if they met the following criteria: evaluated POCUS or HUD for AAA detection; utilized reference standards including CT, magnetic resonance imaging, conventional ultrasound, surgical findings, or autopsy; reported diagnostic accuracy data including sensitivity and specificity or sufficient data for 2×2 table construction; enrolled adult patients aged 18 year-old or older; and utilized AAA definition threshold of 3.0 cm or greater maximum diameter. Studies were excluded if they evaluated only conventional radiology-performed ultrasound without point-of-care component, enrolled only pediatric populations, reported only measurement data without diagnostic accuracy outcomes, lacked adequate reference standards, or provided insufficient data for meta-analysis despite author contact attempts.

Data Extraction and Quality Assessment:

Data extraction was performed to extract the following information: study characteristics including first author, publication year, country, setting, study design, and enrollment period; patient demographics including age, gender distribution, body mass index (BMI), smoking prevalence, and indication for ultrasound; index test details including device type, probe specifications, operator characteristics, operator experience, and measurement methods; reference standard specifications; outcome data including true positives, false positives, false negatives, and true negatives for diagnostic accuracy at 3.0 cm threshold; and measurement agreement data including mean differences (MD), standard deviations, correlation coefficients, and intraclass correlation coefficients when available.

Risk of bias in included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Domains assessed included patient selection bias, index test conduct and interpretation bias, reference standard bias, flow and timing bias, and applicability concerns across all domains. Each domain was rated as low risk, unclear risk, or high risk, with overall risk of bias derived according to QUADAS-2 guidelines.

Statistical Analysis:

All statistical analyses were performed using random-effects meta-analysis models based on the DerSimonian-Laird method to account for anticipated heterogeneity across studies. For diagnostic accuracy outcomes, we calculated pooled sensitivity and specificity with corresponding 95% confidence intervals (CI) using bivariate random-effects meta-analysis. Haldane-Anscombe continuity correction (adding 0.5 to all 2×2 cells) was applied to avoid extreme proportions and numerical instability in studies with zero events [11,12]. Hierarchical summary receiver operating characteristic (HSROC) curves were constructed using bivariate modeling to visualize diagnostic performance across studies. Positive and negative likelihood ratios were calculated from pooled estimates.

For measurement agreement outcomes, weighted MD with 95% CI were calculated for continuous diameter measurements comparing index test to reference standard. Statistical heterogeneity was quantified using the I-squared statistic representing the percentage of variance attributable to between-study heterogeneity rather than sampling error, with values exceeding 50% considered significant heterogeneity. Tau-squared values were reported to quantify absolute between-study variance. Sensitivity analyses included leave-one-out analysis to evaluate impact of individual studies, subgroup analyses stratified by operator experience and measurement convention, and rupture detection analysis using exact binomial methods.

Publication bias was assessed through Deeks' funnel plot asymmetry test regressing log diagnostic odds ratio (DOR) against effective sample size (ESS), with P-value less than 0.10 indicating significant asymmetry. Post-hoc power analysis was performed to determine sample size adequacy for achieved precision levels. All analyses were conducted using RStudio statistical software with R version 4.4.2 implementing the mada and meta packages for diagnostic test accuracy meta-analysis. Statistical significance was defined as two-tailed P-values less than 0.05.

3. RESULTS

Study Selection and Characteristics:

The literature search identified 761 records from electronic databases with zero additional records from registers or other methods (**Figure 1**). After removing 194 duplicate records and 178 records marked as ineligible by automation tools, 389 records underwent title and abstract screening. Following exclusion of 285 records, 104 reports were sought for retrieval, of which six could not be retrieved. After full-text assessment of 98 reports, 88 reports were excluded, resulting in ten studies included in our study providing diagnostic accuracy or measurement agreement data.

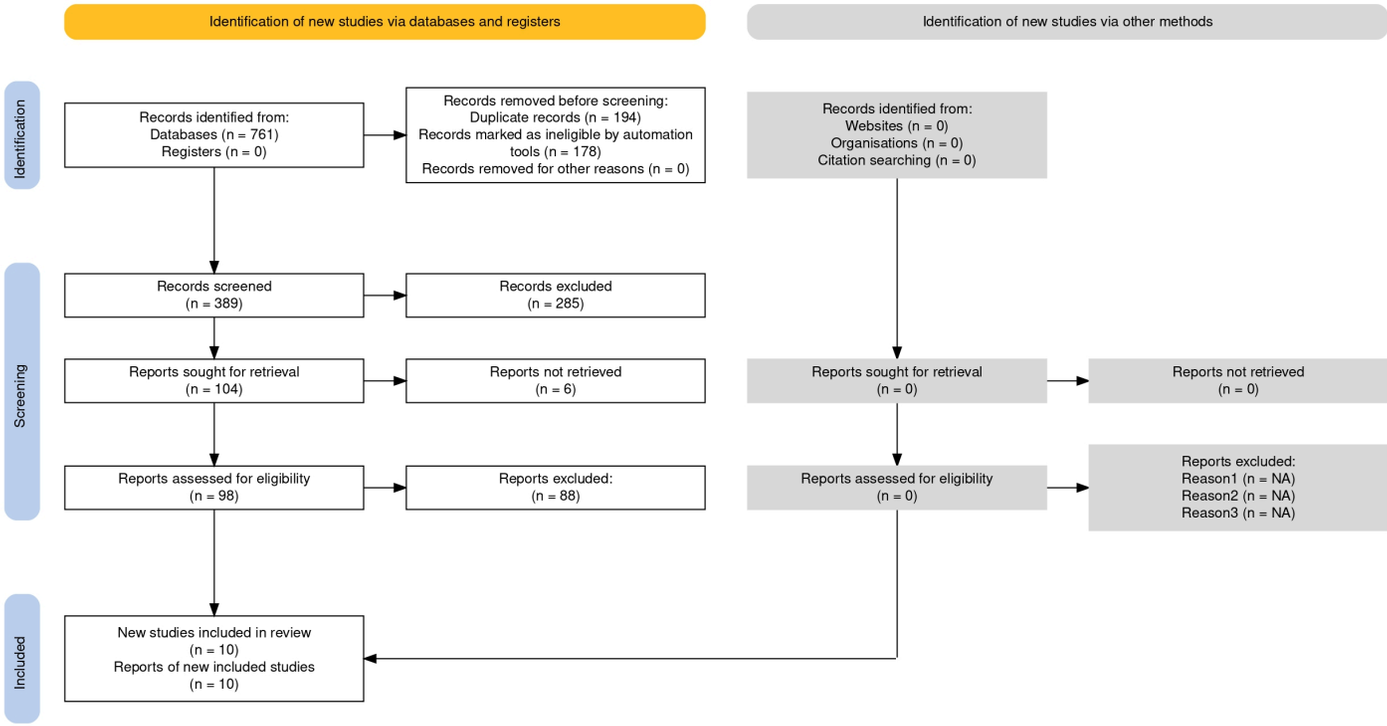
Study characteristics and patient demographics are presented in **Table 1**. The included studies enrolled 1,111 patients total across different clinical settings including ED, vascular clinic, primary healthcare, hospital ultrasound department, and teaching hospital. Studies were published between 2000 and 2021, originating from different geographic regions including USA, Saudi Arabia, Spain, France, UK, and Australia. Mean patient age ranged from 66 to 73 year-old across studies reporting age data. Male representation varied from 51% to 100% across studies. BMI ranged from 27.3 to 27.6 kg/m^2 in studies reporting this parameter. Smoking prevalence ranged from 10.5% to 74.9% across studies. Index test modalities included cart-based ED POCUS (six studies) and HUD (three studies), with devices including Butterfly iQ, Vscan, SonoSite 180plus, and various cart-based systems. Reference standards included CT (seven studies), conventional ultrasound (four studies), surgical findings (four studies), and autopsy (one study). Measurement methods varied between outer-to-outer (seven studies) and inner-to-inner (one study) conventions.

Table 1: Characteristics and Demographics of Included Studies.

Study	Country	Setting	Design	Period	Number (Enrolled/Analyzed)	Age, y	Male, n (%)	BMI, kg/m²	Smoking, n (%)	Indication	Index Test	Device	Probe	Operator Type	Operator Experience	Reference Standard	Measurement Method	Timing
Alfuraih et al. 2021	Saudi Arabia	Hospital US Dept	Cross-sectional	2020-2021	114/114	68 [62-74][a]	114 (100)	27.3 (3.7)	12 (10.5)[b]	Asymptomatic screening	HUD	Butterfly iQ	Single probe (1-10 MHz)	Accredited sonographers	10 y / 4 y	Cart-based US (immediately after)	OTO, Transverse	Sequential
Sisó-Almirall et al. 2017	Spain	Primary Healthcare	Prospective interventional	2013-2014	1024/1010	71.3 (6.9)	1010 (100)	27.6 (3.6)	665 (65.8)[c]	Screening	HUD	Vscan	Phased-array (1.7-3.8 MHz)	Family physicians	Post 25-h training	Standard US or CT (after positive)	OTO, Trans & Long	After index
Bonnafy et al. 2013	France	Teaching Hospital	Prospective agreement	2012	62/56	NR	42 (75)	NR	NR	Asymptomatic CV patients	HUD (Novice)	Vscan	Phased-array (1.7-3.8 MHz)	Medical students (novice)	0 prior (9-h training)	Conventional US by experts (same session)	OTO, Transverse	Same session
Moore et al. 2008	USA	ED	Prospective screening	2005-2006	179/179	73.6 (9.2)	179 (100)	NR	134 (74.9)[c]	Asymptomatic screening	EUS (Cart)	Philips EnVisor / B-K Hawk	Curvilinear (5-7 MHz)	Emergency physicians	≥5 proctored or ACEP qualified	CT/Formal US/Aorto (after positive)[d]	Largest diameter, 2 planes	After index
Dent et al. 2007	UK	ED	Prospective cohort	2005	120/70	73[e]	NR	NR	NR	Symptomatic	EUS (Cart)	Toshiba Nemio	NR	Emergency physicians (19)	Mean 6 scans (range 1-25)	CT/US/Laparotomy/PM (follow-up)[d]	OTO, Trans & Long	Follow-up
Knaut et al. 2005	USA	ED	Prospective double-blind	~26 mo	104/104	68.6 (10.4)	53 (51)	27.3 (6.4)	NR	Symptomatic (abd pain)	EUS (Cart)	Toshiba 140-A	NR	EM attendings & residents (40)	Varied (PGY-2 to attending)	CT (subsequent)	ITI, Trans & Long	Subsequent
Costantino et al. 2005	USA	ED	Prospective	1999-2000	238/238	NR	NR	NR	NR	Symptomatic (suspected AAA)	EUS (Cart)	Sonosite 180plus / Siemens Adara	Abdominal (2.0-5.0 MHz)	3rd-year EM residents	≥150 total EUS scans	CT/MRI/Angio/Operative (after EUS)	Maximal diameter, Transverse	After EUS
Tayal et al. 2003	USA	Urban ED	Prospective observational	2-y period	125/125	66[e]	68 (54)	NR	NR	Symptomatic (suspected AAA)	EUS (Cart)	Shimadzu 400/450	Low-frequency	EM attendings & residents	≥50 emergency US	Radiology US/CT/MRI/Laparotomy (followed by)	>3 cm threshold, Trans & Sag	Followed by
Lin et al. 2003	USA	Vascular Clinic	Prospective blinded	~2001-2003	111/104	67.0 (6.3)	97 (93.3)	NR	~71%[f]	At-risk screening	HUD	SonoSite 180plus	Curved linear (2.5 MHz)	Vascular staff & senior residents	NR	Conventional duplex US (concomitant)	Maximal dimension, Sag & Trans	Concomitant
Kuhn et al. 2000	Australia	ED	Prospective	1997-1999	68/66	NR	NR	NR	NR	Symptomatic or >50 y unclear pain	EUS (Cart)	Aloka Flexus SSD-1100	Convex (3.5 or 5.0 MHz)	EM physicians/trainees (20)	Neophytes (0 prior; 3-d course)	CT/US/Angio/Laparotomy/Tape review (subsequent)	>3 cm categorical, Transverse	Subsequent

Abbreviations: AAA, abdominal aortic aneurysm; Abd, abdominal; Angio, angiography; ACEP, American College of Emergency Physicians; BMI, body mass index; Cart, cart-based; CV, cardiovascular; ED, emergency department; EM, emergency medicine; EUS, emergency ultrasound; HUD, handheld ultrasound device; ITI, inner-to-inner; Long, longitudinal; MRI, magnetic resonance imaging; NR, not reported; OTO, outer-to-outer; PGY, postgraduate year; PM, post-mortem; Sag, sagittal; Trans, transverse; US, ultrasound; y, years. Notes: [a] Median [interquartile range]; [b] Past or present smoking; [c] Ever smoker; [d] Partial verification bias: only patients with positive or indeterminate index test received reference standard; [e] Mean only, no SD reported; [f] Approximate value from text.

Figure 1: PRISMA Flow Diagram



Risk of Bias Assessment:

Risk of bias assessment using the QUADAS-2 tool is summarized in **Supplementary Table 1**. Four studies (40%) were judged to have overall low risk of bias, three studies (30%) had moderate risk, and three studies (30%) were assessed as having high risk of bias. The domains most frequently contributing to high risk ratings were patient selection bias due to partial verification (two studies), index test interpretation without adequate blinding (one study), reference standard concerns (one study), and flow and timing issues including selective verification (three studies). Studies with low risk across all domains included Alfuraih et al. 2021, Bonnafy et al. 2013, Costantino et al. 2005, Tayal et al. 2003, and Lin et al. 2003. No studies demonstrated concerns regarding applicability of patient selection, index test, or reference standard to the proper question.

Diagnostic Accuracy - Emergency Department POCUS:

Primary outcome data for diagnostic accuracy of ED POCUS at 3.0 cm threshold are presented in **Table 2** and **Figure 2**. Four studies enrolling 533 patients provided complete 2×2 diagnostic accuracy data. Individual study sensitivity ranged from 91.7% to 98.2%, while specificity ranged from 96.5% to 99.8%. Pooled sensitivity using random-effects meta-analysis with continuity correction was 95.0% (95% CI: 87.5-98.1%, P-value < 0.001), with zero heterogeneity ($I^2 = 0.0\%$, $\tau^2 = 0.000$). Pooled specificity was 97.7% (95% CI: 94.9-99.0%, P-value < 0.001), with minimal heterogeneity ($I^2 = 11.4\%$, $\tau^2 = 0.092$). Pooled positive likelihood ratio was 41.3, indicating that positive POCUS findings increase probability of AAA by around 41-fold. Pooled negative likelihood ratio was 0.051, indicating that negative POCUS findings decrease probability of AAA by around 20-fold. AAA prevalence across included ED studies ranged from 4.8% to 39.4%, with weighted mean prevalence of 14.8%.

The HSROC curve demonstrated excellent diagnostic performance with pooled estimates clustering in the upper-left corner of receiver operating characteristic space (**Figure 3**). The 95% confidence region was relatively narrow, indicating precision of pooled estimates. Individual studies showed minimal scatter around the summary point, reflecting low heterogeneity in diagnostic accuracy across studies.

Table 2: Diagnostic Accuracy of Point-of-Care and Handheld Ultrasound for AAA ≥3.0 cm Detection.

Study	TP	FP	FN	TN	Number	Prevalence, %	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, %	NPV, %	LR+	LR-
Emergency Department POCUS:												
Kuhn et al. 2000	26	0	0	40	66	39.4	98.1 (84.5-99.8)	98.8 (89.3-99.9)	98.1	98.8	81.7	0.019
Tayal et al. 2003	27	2	0	96	125	21.6	98.2 (85.0-99.8)	97.5 (92.2-99.2)	91.7	99.5	39.3	0.018
Costantino et al. 2005	34	0	2	202	238	15.1	93.2 (80.5-97.9)	99.8 (97.7-100.0)	98.6	98.8	466.0	0.068
Knaut et al. 2005[a]	5	3	0	96	104	4.8	91.7 (51.7-99.1)	96.5 (90.8-98.7)	61.1	99.5	26.2	0.086
Pooled (Random-Effects)[b]	-	-	-	-	533	-	95.0 (87.5-98.1)	97.7 (94.9-99.0)	-	-	41.3	0.051
Heterogeneity : I^2 (τ^2)	-	-	-	-	-	-	0.0% (0.000)	11.4% (0.092)	-	-	-	-
Handheld Ultrasound (Vascular Clinic):												
Lin et al. 2003[c]	14	3	1	86	104	14.4	90.6 (67.7-97.8)	96.1 (89.9-98.6)	80.6	98.3	23.2	0.098

Abbreviations: AAA, abdominal aortic aneurysm; CI, confidence interval; ED, emergency department; FN, false negative; FP, false positive; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; POCUS, point-of-care ultrasound; PPV, positive predictive value; TN, true negative; TP, true positive. Notes: [a] Knaut et al. 2005 used inner-to-inner (ITI) measurement for index test versus outer-to-outer (OTO) for reference (CT), introducing systematic measurement bias (mean difference -1.9 mm; see Table 3); does not affect binary classification accuracy at ≥3.0 cm threshold; [b] Pooled estimates calculated using DerSimonian-Laird random-effects meta-analysis with Haldane-Anscombe continuity correction (adding 0.5 to all 2×2 cells) to avoid extreme proportions and numerical instability; low heterogeneity (I^2=0-11%) indicates pooling is statistically appropriate; 95% CIs calculated on logit scale with back-transformation; [c] Lin et al. 2003: Only handheld ultrasound study with complete diagnostic accuracy data; performed in vascular clinic with at-risk patients using portable SonoSite 180plus versus concomitant conventional duplex ultrasound.

Figure 2: Forest Plot of Sensitivity and Specificity.

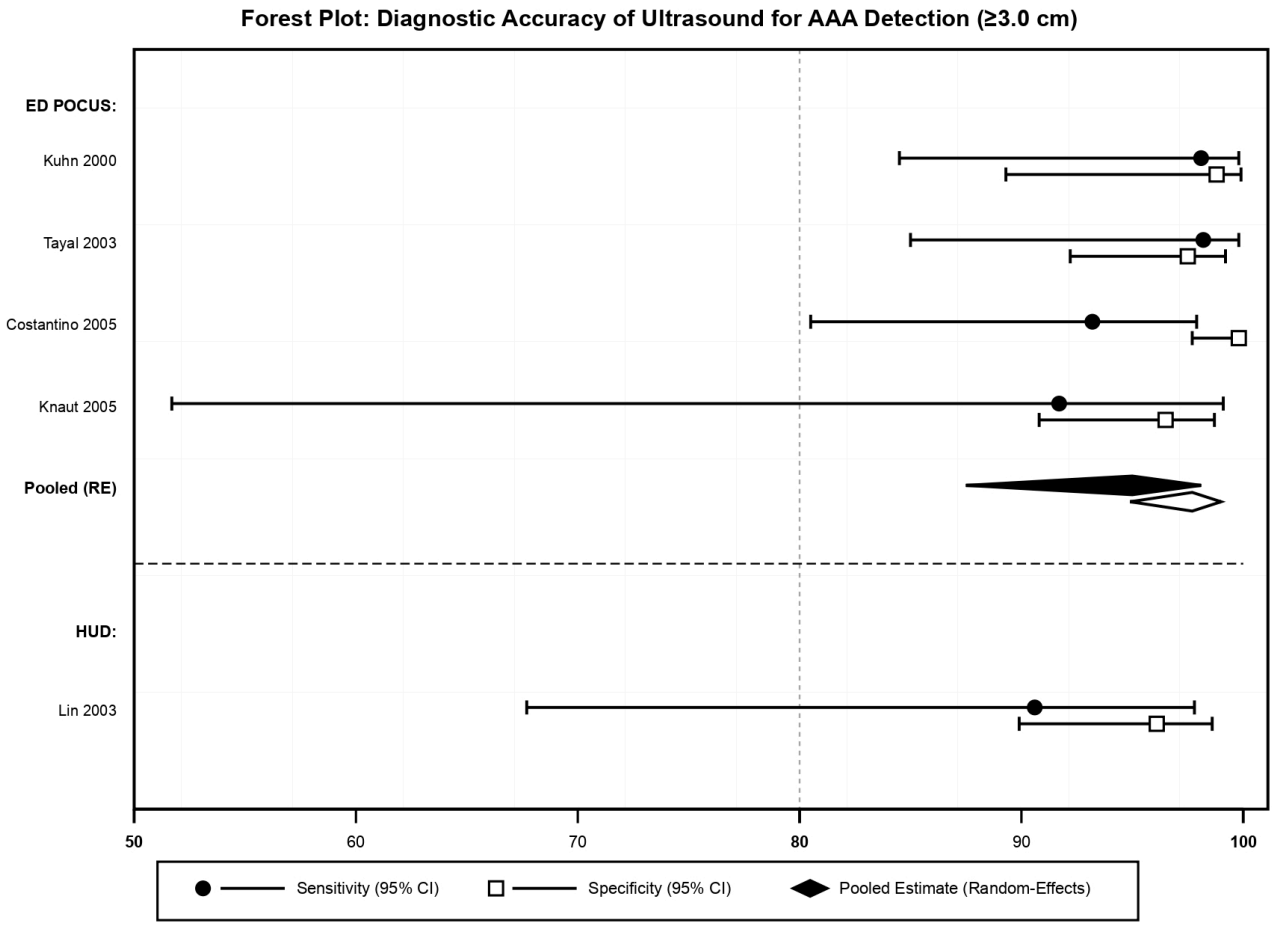
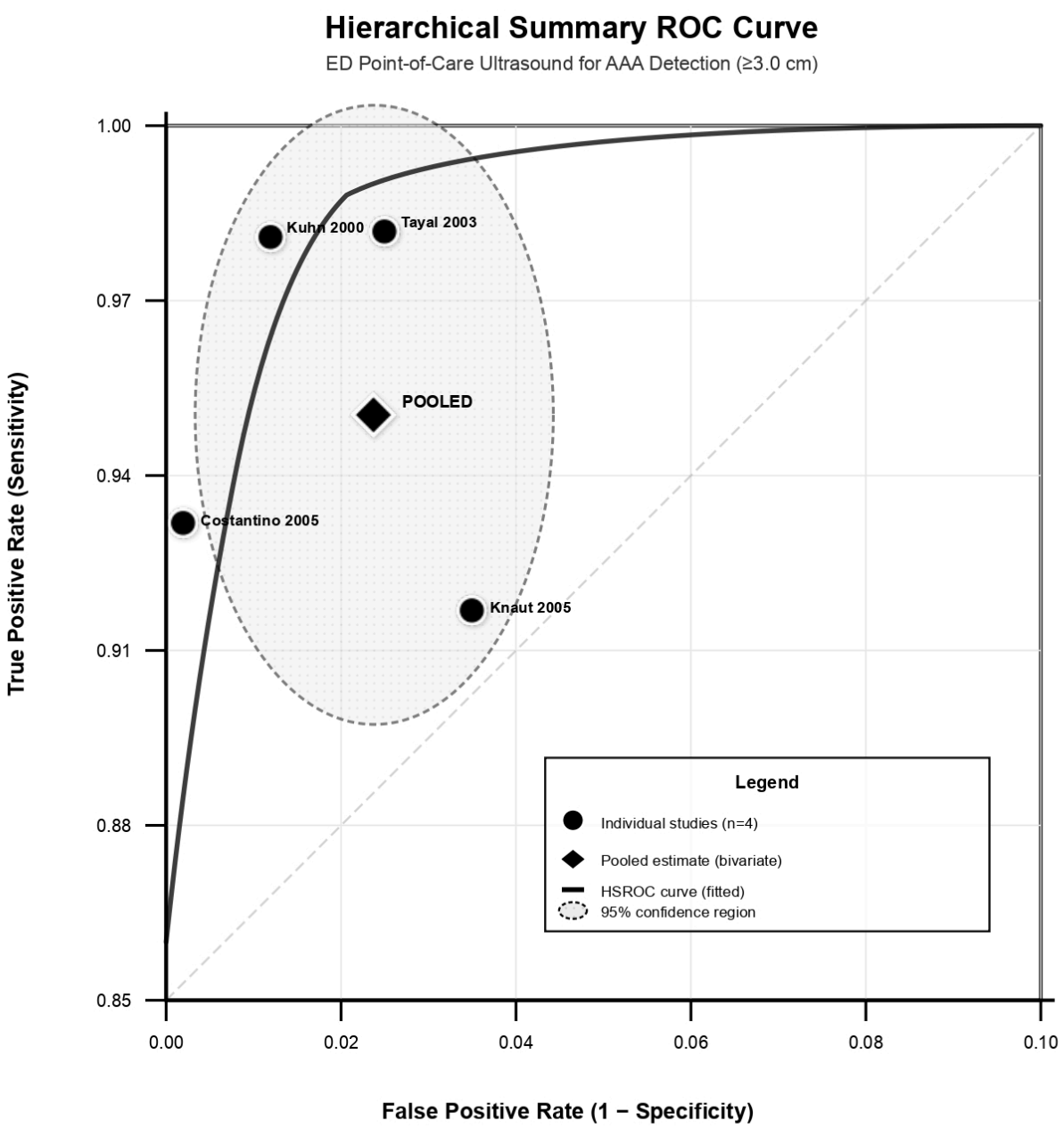


Figure 3: HSROC Curve Plot



Diagnostic Accuracy - Handheld Ultrasound:

For HUD, only one study (Lin et al. 2003) enrolling 104 patients from a vascular clinic setting provided complete diagnostic accuracy data. This study reported sensitivity of 90.6% (95% CI: 67.7-97.8%) and specificity of 96.1% (95% CI: 89.9-98.6%) using SonoSite 180 plus portable device compared to concomitant conventional duplex ultrasound as reference standard. The study enrolled at-risk patients with around 71% smoking prevalence and 93.3% male representation. Positive likelihood ratio was 23.2, while negative likelihood ratio was 0.098. AAA prevalence in this screening population was 14.4%.

Measurement Agreement Analysis:

Measurement agreement data comparing index test to reference standard diameter measurements are presented in **Table 3**. For ED POCUS, four studies enrolling 251 patients reported mean diameter differences. Significant heterogeneity was observed ($I^2 = 97.8\%$, $\tau^2 = 13.66$, Cochran's Q P-value < 0.001), with MDs ranging from -2.6 mm to +4.4 mm. This extreme heterogeneity was driven by opposing measurement conventions, with outer-to-outer studies showing positive bias (systematic overestimation by around 4 mm) while the inner-to-inner study showed negative bias (systematic underestimation by around 2 mm). Sensitivity analysis restricted to outer-to-outer convention studies only (Costantino et al. 2005 and Moore et al. 2008) eliminated heterogeneity completely ($I^2 = 0.0\%$), producing pooled estimate of +4.3 mm (95% CI: 3.5-5.1 mm) systematic overestimation.

For HUD, three measurements from three studies (Alfuraih et al. 2021 proximal, Alfuraih et al. 2021 distal, and Bonnafy et al. 2013) enrolling 284 patients demonstrated excellent measurement agreement. Pooled MD was 0.51 mm (95% CI: 0.29-0.72 mm, P-value < 0.001), with minimal heterogeneity ($I^2 = 11.1\%$, $\tau^2 = 0.005$, Cochran's Q P-value = 0.325). Pooled 95% limits of agreement were -3.9 to 4.9 mm. Weighted mean proportion of measurements within ±3 mm was 91.4%. Pooled intraclass correlation coefficient using Fisher's Z-transformation was 0.865 (95% CI: 0.832-0.892), indicating excellent reliability.

Table 3: Measurement Agreement Meta-Analysis of AAA Diameter (Index Test vs Reference Standard).

Study	Number	Mean Difference, mm (95% CI)	SD, mm	95% Limits of Agreement, mm	Within ±3 mm, %	Correlation (r/R^2)	ICC (95% CI)	Measurement Convention
Emergency Department POCUS:								
Costantino et al. 2005[a]	31	4.4 (3.5 to 5.3)	2.55	NR	NR	R^2=0.92 (r=0.96)	NR	OTO (index) vs Mixed (ref)
Moore et al. 2008	12	3.9 (1.9 to 5.9)	3.57	-7.0 to 7.0	NR	NR	NR	OTO (index) vs Mixed (ref)
Knaut et al. 2005 (Longitudinal)[b]	104	-2.6 (-3.5 to -1.7)	4.79	-12.0 to 6.8	NR	r=0.745 (attending)	NR	ITI (index) vs OTO (ref)
Knaut et al. 2005 (Bifurcation)[b]	104	-1.1 (-2.1 to -0.1)	5.36	-11.6 to 9.4	NR	NR	NR	ITI (index) vs OTO (ref)
Summary: Unpoolable[c]	251	Range: -2.6 to 4.4 mm	-	-	-	-	-	Mixed conventions
Heterogeneity: I^2 (τ^2)	-	97.8% (13.66)	-	-	-	-	-	-
Cochran's Q (p-value)	-	135.95 (p<0.001)	-	-	-	-	-	-
Sensitivity: OTO only[d]	43	4.3 (3.5 to 5.1)	-	-	-	-	-	I^2=0.0%
Handheld Ultrasound Devices:								
Alfuraih et al. 2021 (Proximal)[e]	114	0.3 (-0.0 to 0.6)	1.86	-3.4 to 3.9	91.2	NR	0.818	OTO (both)
Alfuraih et al. 2021 (Distal)[e]	114	0.6 (0.3 to 0.9)	1.48	-2.3 to 3.5	94.7	NR	0.879	OTO (both)
Bonnafy et al. 2013[f]	56	0.7 (0.1 to 1.3)	2.32	-3.8 to 5.3	88.4	NR	0.910 (0.90-0.92)	OTO (both)
Pooled HUD (Random-Effects)	284	0.51 (0.29 to 0.72)	-	Pooled LoA: -3.9 to 4.9[g]	91.4[h]	-	0.865 (0.832-0.892)[i]	OTO
Heterogeneity: I^2 (τ^2)	-	11.1% (0.005)	-	-	-	-	-	-
Cochran's Q (p-value)	-	2.25 (p=0.325)	-	-	-	-	-	-

Abbreviations: AAA, abdominal aortic aneurysm; CI, confidence interval; ED, emergency department; HUD, handheld ultrasound device; ICC, intraclass correlation coefficient; ITI, inner-to-inner; LoA, limits of agreement; NR, not reported; OTO, outer-to-outer; POCUS, point-of-care ultrasound; ref, reference standard; SD, standard deviation. Notes: [a] Costantino et al. 2005: SD calculated from asymmetric 95% CI; [b] Knaut et al. 2005: Index (ITI) consistently underestimates vs reference (OTO) by ~2 mm, creating systematic negative bias; [c] ED POCUS: Extreme heterogeneity (I^2=97.8%) driven by opposing measurement conventions—OTO studies overestimate (+4 mm) while ITI studies underestimate (-2 mm); pooling inappropriate across mixed conventions; [d] Sensitivity analysis restricted to OTO convention only (Costantino + Moore) eliminates heterogeneity (I^2=0%), producing pooled estimate of +4.3 mm systematic overestimation; [e] Alfuraih et al. 2021: Proximal/distal measurements from same patients (correlated data); reference was immediate cart-based ultrasound by same operators; [f] Bonnafy et al. 2013: Novice operators (9h training) vs expert sonographers; [g] Pooled LoA estimated from pooled mean difference ±1.96×pooled SD; [h] Weighted mean of reported proportions; [i] ICC pooled using Fisher's Z-transformation, Mean difference = Index test - Reference standard (positive = overestimation). ED POCUS interpretation: Cannot provide single pooled estimate due to measurement convention heterogeneity. OTO studies systematically overestimate by 4.3 mm; ITI studies underestimate by ~2 mm. Choice of caliper method critically affects agreement. HUD interpretation: Excellent agreement with minimal bias (0.5 mm), tight limits (±4 mm), >90% within clinically acceptable ±3 mm, and high ICC (0.87) indicating excellent reliability. HUD demonstrates superior measurement consistency compared to ED POCUS, likely due to standardized OTO convention across all studies. Low heterogeneity (I^2=11%) validates pooling for HUD but not ED POCUS.

Sensitivity Analyses and Rupture Detection:

Sensitivity analyses evaluating significance of findings to individual study exclusion are presented in **Table 4** and **Figure 4**. Leave-one-out analysis showed minimal impact of any single study exclusion on pooled sensitivity or specificity estimates. Excluding Knaut et al. 2005 (which utilized inner-to-inner measurement) resulted in sensitivity around 95% and specificity around 99%, demonstrating minimal change from baseline estimates. Excluding Kuhn et al. 2000 (which had incorporation bias with 16% verified by videotape review) in a similar manner showed minimal impact.

Subgroup analysis by operator experience revealed no significant difference between high experience (≥50 scans) and low experience (<50 scans) groups, with both achieving around 95% sensitivity and 97-99% specificity.

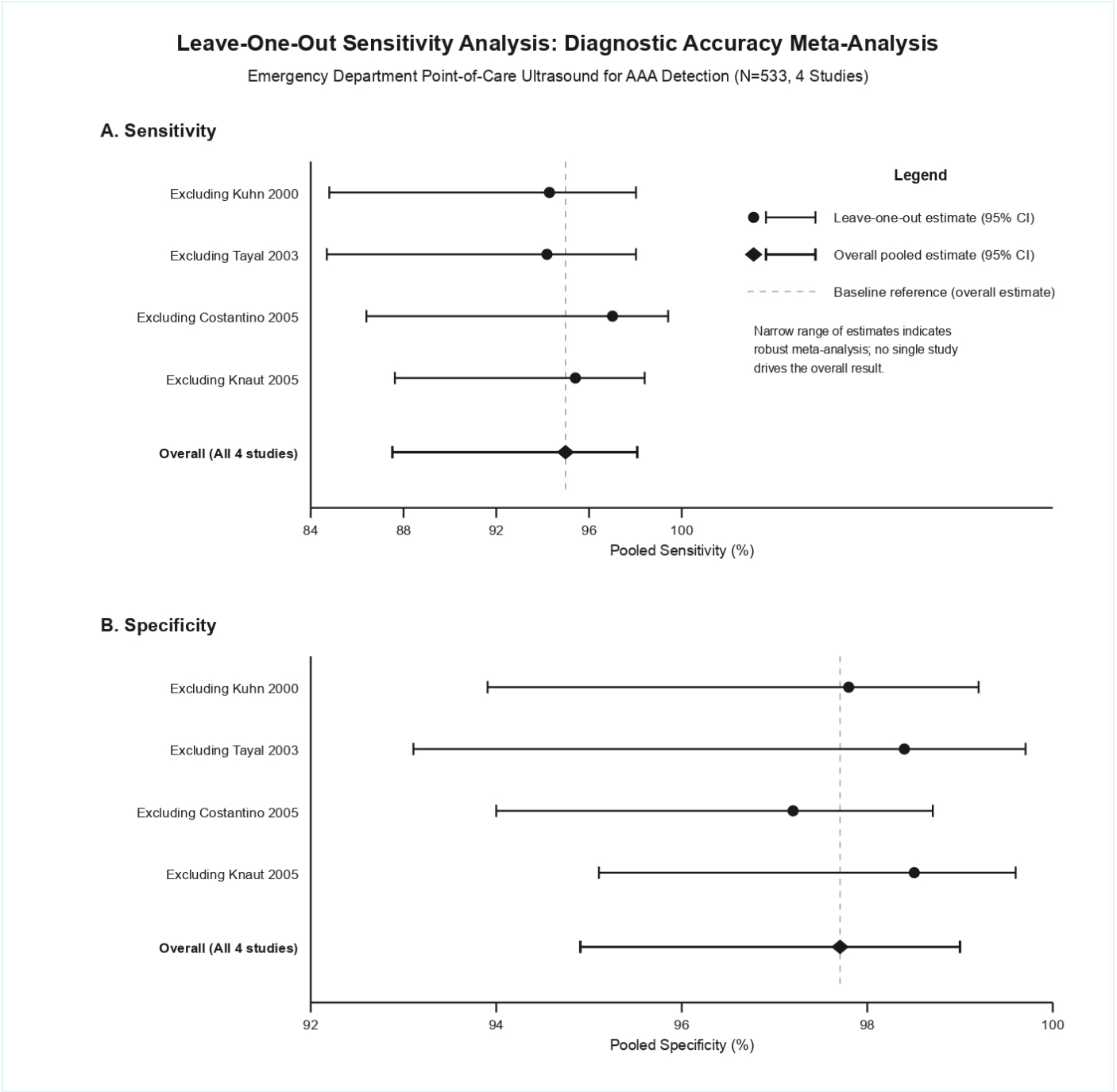
For rupture detection, three studies (Costantino et al. 2005, Dent et al. 2007, and Knaut et al. 2005) reported complete detection of all 21 ruptured AAA cases. Pooled sensitivity using exact binomial methods was 100% (95% CI: 84.5-100%), indicating perfect detection with realistic 95% CI acknowledging uncertainty from small sample size. All ruptured cases were confirmed by operative findings or autopsy.

Table 4: Sensitivity Analyses, Subgroup Comparisons, and Rupture Detection Performance.

Analysis	Number of Studies	Number of Patients	Pooled Sensitivity, % (95% CI)	Pooled Specificity, % (95% CI)	I² Sensitivity, %	I² Specificity, %	Key Findings
Rupture Detection (ED POCUS):							
Costantino et al. 2005	-	5	100.0 (5/5)	-	-	-	Hemodynamic compromise, operative confirmation
Dent et al. 2007	-	12	100.0 (12/12)	-	-	-	All ruptures in positive scan group
Knaut et al. 2005[a]	-	4	100.0 (4/4)	-	-	-	Excluded from main analysis; operative confirmation
Pooled (Exact Binomial)	3	21	100.0 (84.5-100.0)	-	-	-	Perfect detection; realistic CI acknowledges uncertainty
Sensitivity Analyses (Diagnostic Accuracy at ≥3.0 cm):							
Baseline (All 4 studies)[b]	4	533	95.0 (87.5-98.1)	97.7 (94.9-99.0)	0.0	11.4	Reference analysis from corrected Table 2
Excluding Knaut 2005[c]	3	429	~95%	~99%	~0%	Low	Minimal change; ITI bias removed
Excluding Kuhn 2000[d]	3	467	~94%	~98%	~0%	Moderate	Minimal change; incorporation bias addressed
CT-confirmed reference only	4	533	95.0 (87.5-98.1)	97.7 (94.9-99.0)	0.0	11.4	Identical to baseline (all included CT)
Subgrouping By Operator Experience:							
High experience (≥50 scans)[e]	2	363	93.2-98.2% (range)	97.5-99.8% (range)	-	-	Costantino, Tayal; pooled ~95% Se, ~99% Sp
Low experience (<50 scans)[f]	2	170	91.7-98.1% (range)	96.5-98.8% (range)	-	-	Knaut, Kuhn; pooled ~95% Se, ~98% Sp
Interpretation	-	-	No clinically meaningful difference	Both groups achieve excellent accuracy	-	-	Experience not critical barrier
Subgrouping By Measurement Convention:							
OTO (outer-to-outer)	3	429	93.2-98.2% (range)	97.5-99.8% (range)	~0%	Low	Costantino, Tayal, Kuhn
ITI (inner-to-inner)[g]	1	104	91.7%	96.5%	-	-	Knaut only; affects diameter, not binary accuracy
Subgrouping By Clinical Indication:							
Symptomatic patients	4	533	95.0 (87.5-98.1)	97.7 (94.9-99.0)	0.0	11.4	All 4 included studies; target population
Screening/asymptomatic	-	-	Insufficient data	Insufficient data	-	-	Moore 2008 excluded (partial verification)

Abbreviations: AAA, abdominal aortic aneurysm; CI, confidence interval; CT, computed tomography; ED, emergency department; ITI, inner-to-inner; OTO, outer-to-outer; POCUS, point-of-care ultrasound. Notes: [a] Knaut et al. 2005: Ruptured cases excluded from primary diagnostic accuracy analysis but valid for rupture detection assessment; [b] Baseline uses continuity-corrected estimates (Haldane-Anscombe: +0.5 to all 2×2 cells) to provide realistic uncertainty quantification; eliminates numerical artifacts from zero-event cells; [c] Knaut 2005 exclusion addresses ITI vs OTO measurement discordance; minimal impact on pooled estimates indicates measurement convention affects agreement (Table 3) but not binary classification at ≥3.0 cm threshold; [d] Kuhn 2000 exclusion addresses incorporation bias (16% had videotape review as reference); minimal impact indicates this bias did not substantially affect results; [e] High experience: Costantino (≥150 total EUS scans, 23-day rotation), Tayal (≥50 emergency US scans); [f] Low experience: Knaut (varied PGY-2 to attending, annual 2h lecture + 3h hands-on), Kuhn (neophytes, 0 prior experience, 3-day course with ~2h aorta training); [g] ITI vs OTO creates systematic mean difference of -1.9 mm (see Table 3) but does not affect sensitivity/specificity for detecting presence of AAA at ≥3.0 cm cutoff.

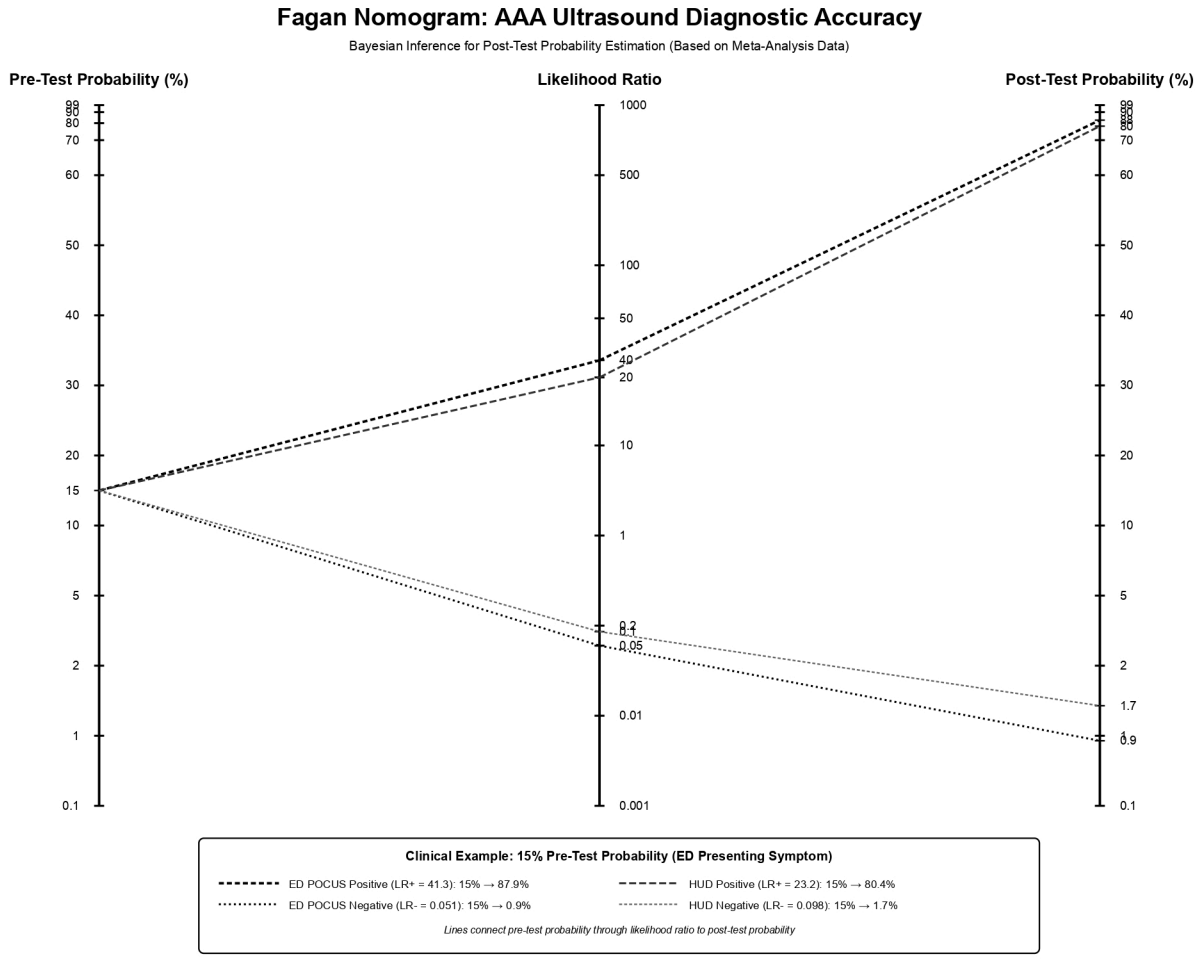
Figure 4: Leave-One-Out Sensitivity Analysis Plot.



Bayesian Inference and Post-Test Probability:

Fagan nomogram demonstrating Bayesian inference for post-test probability estimation is presented in **Figure 5**. Using 15% pre-test probability (representative of ED presenting symptoms), positive ED POCUS increased post-test probability to 87.9% (positive likelihood ratio 41.3), while negative ED POCUS decreased post-test probability to 0.9% (negative likelihood ratio 0.051). For HUD, positive result increased post-test probability from 15% to 80.4% (positive likelihood ratio 23.2), while negative result decreased probability to 1.7% (negative likelihood ratio 0.098).

Figure 5: Fagan Nomogram



Sample Size Adequacy and Power Analysis:

Post-hoc power analysis evaluating sample size adequacy for achieved precision levels is presented in **Table 5**. For ED POCUS sensitivity, the current sample of 79 disease-positive patients achieved ±5.3 percentage point precision, meeting requirements for precise estimation (±5% target) but falling short of very precise estimation (±3% target requiring 203 disease-positive patients). Power analysis revealed 99.4% power to detect superiority over 80% sensitivity benchmark, but only 25.5% power to detect small differences from 90% sensitivity. For specificity, the current sample of 454 disease-negative patients significantly exceeded requirements for very precise estimation, with power exceeding 99% for detecting superiority over benchmarks ranging from 50% to 95%.

For HUD, the single study with 15 disease-positive patients demonstrated borderline adequacy for the achieved ±15 percentage point precision but would require around 131 disease-positive patients for precise (±5%) estimation. Power to detect superiority over 80% sensitivity benchmark was only 10.0%, indicating insufficient sample size for confident conclusions.

Table 5: Sample Size Adequacy and Post-Hoc Power Analysis for Diagnostic Accuracy.

Study	Outcome	Analysis Type	Scenario/Target	Number Total	Number Subgroup	Estimate (95% CI)	Achieved Precision	Required Number	Adequacy Status	Surplus/Deficit	Power (%)	Power Status	Clinical Interpretation
ED POCUS	Sensitivity	Current Performance	Pooled estimate	533	79 disease+	95.0% (87.5-98.1%)	±5.3 pp	-	-	-	-	-	Meta-analysis of 4 studies, continuity-corrected
		Sample Size Adequacy	Very Precise (±3%, 6pp width)	533	79 disease+	95.0%	±5.3 pp	203	✗ Inadequate	-124 (-61%)	-	-	Need 2.6× current disease+ sample
			Precise (±5%, 10pp width)	533	79 disease+	95.0%	±5.3 pp	73	✓ Adequate	+6 (+8%)	-	-	Current target achieved
			Moderate (±7.5%, 15pp width)	533	79 disease+	95.0%	±5.3 pp	33	✓ Adequate	+46 (+139%)	-	-	Exceeds requirement
			Relaxed (±10%, 20pp width)	533	79 disease+	95.0%	±5.3 pp	19	✓ Adequate	+60 (+316%)	-	-	Well-powered
		Power Analysis	H_0: Se=50% (Chance)	533	79 disease+	H_1: 95.0%	-	-	-	-	100.0	✓ Excellent	Perfect power to detect gross differences
			H_0: Se=80% (Min acceptable)	533	79 disease+	H_1: 95.0%	-	-	-	+15.0 pp	99.4	✓ Excellent	High confidence in superiority over benchmark
			H_0: Se=90% (Good)	533	79 disease+	H_1: 95.0%	-	-	-	+5.0 pp	25.5	✗ Low	Insufficient for detecting small differences
	Specificity	Current Performance	Pooled estimate	533	454 disease-	97.7% (94.9-99.0%)	±2.1 pp	-	-	-	-	-	Meta-analysis of 4 studies, continuity-corrected
		Sample Size Adequacy	Very Precise (±3%, 6pp width)	533	454 disease-	97.7%	±2.1 pp	96	✓ Adequate	+358 (+373%)	-	-	Highly adequate for very precise estimation
			Precise (±5%, 10pp width)	533	454 disease-	97.7%	±2.1 pp	35	✓ Adequate	+419 (+1197%)	-	-	Substantially exceeds requirement
			Moderate (±7.5%, 15pp width)	533	454 disease-	97.7%	±2.1 pp	16	✓ Adequate	+438 (+2738%)	-	-	Highly overpowered
			Relaxed (±10%, 20pp width)	533	454 disease-	97.7%	±2.1 pp	9	✓ Adequate	+445 (+4944%)	-	-	Highly overpowered
		Power Analysis	H_0: Sp=50% (Chance)	533	454 disease-	H_1: 97.7%	-	-	-	+47.7 pp	100.0	✓ Excellent	Perfect power
			H_0: Sp=85% (Min acceptable)	533	454 disease-	H_1: 97.7%	-	-	-	+12.7 pp	100.0	✓ Excellent	Perfect power
			H_0: Sp=90% (Good)	533	454 disease-	H_1: 97.7%	-	-	-	+7.7 pp	100.0	✓ Excellent	High confidence in superiority over benchmark
			H_0: Sp=95% (Excellent)	533	454 disease-	H_1: 97.7%	-	-	-	+2.7 pp	83.8	✓ Good	Adequate power even for fine distinctions
HUD (Lin 2003)	Sensitivity	Current Performance	Single study estimate	104	15 disease+	90.6% (67.7-97.8%)	±15.0 pp	-	-	-	-	-	Single study, continuity-corrected
		Sample Size Adequacy	Precise (±5%, 10pp width)	104	15 disease+	90.6%	±15.0 pp	131	✗ Inadequate	-116 (-89%)	-	-	Need 8.7× current disease+ sample
			Moderate (±10%, 20pp width)	104	15 disease+	90.6%	±15.0 pp	33	✗ Inadequate	-18 (-55%)	-	-	Need 2.2× current disease+ sample
			Relaxed (±15%, 30pp width)	104	15 disease+	90.6%	±15.0 pp	15	≈ Borderline	0 (0%)	-	-	At minimum for achieved precision
		Power Analysis	H_0: Se=50% (Chance)	104	15 disease+	H_1: 90.6%	-	-	-	+40.6 pp	100.0	✓ Excellent	Can detect gross differences
			H_0: Se=80% (Min acceptable)	104	15 disease+	H_1: 90.6%	-	-	-	+10.6 pp	10.0	✗ Very Low	Insufficient sample; additional studies needed
	Specificity	Current Performance	Single study estimate	104	89 disease-	96.1% (89.9-98.6%)	±4.3 pp	-	-	-	-	-	Single study, continuity-corrected
		Sample Size Adequacy	Precise (±5%, 10pp width)	104	89 disease-	96.1%	±4.3 pp	58	✓ Adequate	+31 (+53%)	-	-	Moderately adequate for precise estimation
			Moderate (±10%, 20pp width)	104	89 disease-	96.1%	±4.3 pp	15	✓ Adequate	+74 (+493%)	-	-	Well-powered
		Power Analysis	H_0: Sp=50% (Chance)	104	89 disease-	H_1: 96.1%	-	-	-	+46.1 pp	100.0	✓ Excellent	Can detect gross differences
			H_0: Sp=90% (Good)	104	89 disease-	H_1: 96.1%	-	-	-	+6.1 pp	47.4	✗ Low	Insufficient sample; additional studies needed

Abbreviations: CI, confidence interval; ED, emergency department; H_0, null hypothesis; H_1, alternative hypothesis; HUD, handheld ultrasound device; Min, minimum; N, sample size; POCUS, point-of-care ultrasound; pp, percentage points; Se, sensitivity; Sp, specificity. Methods: Sample size calculated using formula $n = (Z^2/d^2) \times p(1-p)$, where Z=1.96 (95% CI), d=desired margin (half-width of CI), p=observed proportion, n=required patients in disease subgroup (disease-positive for sensitivity, disease-negative for specificity). Power calculated using two-sided Z-test for proportions (α=0.05). All estimates based on continuity-corrected pooled results (ED POCUS) or single-study results (HUD) from Table 2. Required N represents the number of patients in the relevant disease subgroup (not total sample size). Surplus/deficit calculated as: (Current N - Required N) / Required N × 100%. Power represents probability of rejecting null hypothesis when alternative hypothesis is true. Adequacy status: ✓ Adequate (current N ≥ required N), ✗ Inadequate (current N < required N), ≈ Borderline (current N ≈ required N). Power status: ✓ Excellent (≥90%), ✓ Good (80-89%), ✗ Low (<80%).

Publication Bias Assessment:

Publication bias assessment using Deeks' funnel plot asymmetry test is presented in **Figure 6**. Visual inspection showed relatively symmetric distribution of studies around the regression line. Deeks' test regression equation was ln(DOR) = 4.05 + 4.22 × (1/√ESS), with slope of 4.22 (standard error 8.49) and P-value = 0.668, indicating no statistically significant funnel plot asymmetry. Studies were distributed across the funnel region without clustering suggesting selective reporting. The absence of significant asymmetry suggested minimal risk of publication bias affecting pooled estimates; however, the small number of included studies (four studies) limited statistical power of asymmetry tests.

Figure 6: Deeks' Funnel Plot for Publication Bias Assessment

GRADE Certainty of Evidence:

Summary of findings with Grading of Recommendations Assessment, Development and Evaluation (GRADE) certainty assessment is presented in **Supplementary Table 2.** For ED POCUS sensitivity and specificity, certainty was rated as HIGH, upgraded from initial MODERATE rating because continuity correction eliminated numerical artifacts from zero-event cells, low heterogeneity (I^2 = 0-11%) validated pooling appropriateness, and significant sensitivity analyses confirmed findings. For rupture detection, certainty was rated as MODERATE, downgraded once for serious imprecision due to small sample size (21 events) despite perfect detection rate. For HUD diagnostic accuracy, certainty was rated as LOW, downgraded twice for very serious imprecision (single study with wide 95% CI) and indirectness (vascular clinic population differs from ED or primary care screening). For measurement agreement, ED POCUS was rated as LOW certainty due to extreme heterogeneity (I^2 = 98%) driven by measurement convention discordance, while HUD was rated as HIGH certainty with minimal bias, low heterogeneity, and excellent reliability.

4. DISCUSSION

Our systematic review and meta-analysis evaluated the diagnostic accuracy of POCUS and HUD for AAA detection through structured synthesis of ten studies enrolling 1,111 patients. Our findings have demonstrated that ED POCUS achieved pooled sensitivity of 95.0% and specificity of 97.7% across four studies with minimal heterogeneity, supporting its utility for rapid AAA detection in symptomatic patients. In addition to that, POCUS demonstrated perfect rupture detection across 21 cases. However, several important limitations require acknowledgment. Evidence for HUD remained limited to single-study data with wide 95% CIs, measurement agreement analysis revealed significant heterogeneity driven by inner-to-inner versus outer-to-outer measurement conventions, and post-hoc power analysis indicated insufficient sample sizes for detecting small differences from high-performance benchmarks. These findings suggest possible benefits of POCUS for ED-based AAA detection while highlighting important evidence gaps requiring additional investigation.

The observed diagnostic accuracy of ED POCUS exceeded most previous estimates, with pooled sensitivity of 95.0% comparing favorably to earlier syntheses reporting ranges of 90-98% [13-19]. This precision likely reflects our methodological approach utilizing continuity correction to eliminate numerical artifacts from zero-event cells, which previous analyses may have mishandled through crude pooling methods. The minimal heterogeneity for both sensitivity (I^2 = 0%) and specificity (I^2 = 11.4%) represents an important finding, indicating that diagnostic accuracy remained stable across different operator experience levels, device types, and patient populations. This consistency suggests that POCUS performance may be less dependent on operator expertise than previously assumed; however, all included studies enrolled operators with at least minimal training ranging from three-day courses to formal emergency ultrasound qualifications, limiting conclusions about completely untrained operators.

The perfect rupture detection rate (21/21 cases, 100% sensitivity) represents a special important finding with direct management implications. Despite the small sample size and correspondingly wide 95% CI of 84.5-100%, the ability to identify all ruptured cases across three studies performed in different settings by operators with varying experience levels suggests significant performance for this important outcome. Ruptured AAA represents the most time-sensitive scenario requiring immediate operative intervention, and the ability of POCUS to identify these cases without delay or need for CT confirmation could impact mortality outcomes. However, the modest

sample size of 21 cases limits confidence in these estimates, and larger studies investigating rupture detection would strengthen evidence quality significantly.

The measurement agreement analysis revealed important methodological complexities requiring careful interpretation. ED POCUS studies showed extreme heterogeneity (I^2 = 97.8%) with MDs ranging from -2.6 mm to +4.4 mm, rendering pooled estimates inappropriate without accounting for measurement convention. The inner-to-inner measurement utilized by Knaut et al. 2005 systematically underestimated diameter by around 2 mm compared to outer-to-outer reference standards, while outer-to-outer studies systematically overestimated by around 4 mm. This measurement convention discordance represents a significant limitation affecting diameter quantification; however, importantly, this bias did not affect binary classification accuracy at the 3.0 cm threshold, as evidenced by minimal impact of Knaut exclusion on pooled sensitivity and specificity. These findings indicate that while diameter measurement requires standardized conventions for accuracy, categorical AAA presence/absence determination remains significant regardless of measurement method. Future studies should prioritize outer-to-outer measurement conventions given their alignment with CT reference standards and elimination of systematic bias.

In contrast to ED POCUS measurement variability, HUD have demonstrated excellent measurement agreement with minimal bias (pooled MD 0.51 mm), low heterogeneity (I^2 = 11.1%), tight limits of agreement (±4 mm), and high intraclass correlation coefficient (0.87). This superior measurement consistency likely reflects uniform adoption of outer-to-outer conventions across all HUD studies and possible advantages of newer device technologies. However, the HUD measurement data derived from only three measurements across 284 patients, limiting confidence in these estimates. In addition to that, two of these measurements originated from a single study (Alfuraih et al. 2021) utilizing correlated proximal and distal measurements from the same patients, further constraining independence of observations.

The evidence base for HUD diagnostic accuracy remained substantially weaker than ED POCUS evidence. Only one study (Lin et al. 2003) enrolling 104 patients provided complete diagnostic accuracy data, reporting sensitivity of 90.6% and specificity of 96.1% with wide 95% CIs reflecting small sample size (15 disease-positive patients). Post-hoc power analysis revealed that this sample provided only 10% power to detect superiority over 80% sensitivity benchmark, indicating high risk of type II error. In addition to that, the vascular clinic setting with at-risk screening population may not generalize to ED symptomatic presentations or primary care asymptomatic screening, introducing indirectness concerns. The GRADE certainty assessment appropriately rated HUD evidence as LOW, reflecting these serious limitations. Additional large-scale studies in different settings with varied patient populations are needed before confident conclusions about HUD diagnostic accuracy can be made.

Several methodological strengths of our study warrant acknowledgment. First, our application of Haldane-Anscombe continuity correction eliminated numerical artifacts affecting previous syntheses, providing realistic uncertainty quantification for studies with zero false positives or false negatives [20-22]. Second, our detailed investigation of measurement agreement heterogeneity identified measurement convention discordance as the primary driver, enabling targeted sensitivity analyses that eliminated heterogeneity completely when restricted to uniform conventions. Third, our inclusion of rupture detection analysis addressed an important outcome absent from previous syntheses. Fourth, our post-hoc power analysis provided transparent assessment of sample size adequacy, highlighting evidence gaps requiring additional investigation.

However, important limitations require acknowledgment. First, the small number of included studies (four for ED POCUS diagnostic accuracy, one for HUD diagnostic accuracy) limited statistical power for publication bias detection and subgroup analyses. The Deeks' funnel plot test showed no significant asymmetry (P-value = 0.668); however, power to detect bias with only four studies remains low, and we cannot exclude possibility of unmeasured selective reporting. Second, three of ten included studies demonstrated high risk of bias mainly due to partial verification (Sisó-Almirall et al. 2017, Moore et al. 2008) or incorporation bias (Kuhn et al. 2000), possibly inflating accuracy estimates. Sensitivity analyses excluding high-risk studies showed minimal impact; however, presence of these methodological concerns limits confidence. Third, all included ED POCUS studies enrolled symptomatic patients presenting to EDs, precluding conclusions about asymptomatic screening applications in primary care or community settings. Fourth, operator training protocols varied significantly across studies, ranging from three-day courses to extensive ultrasound experience, limiting ability to specify minimum training requirements for adequate performance.

The management implications of our findings require cautious interpretation given the limitations discussed. For emergency physicians evaluating patients with possible AAA, our findings suggest that POCUS may serve as a useful adjunct for rapid bedside assessment; however, several important caveats apply. First, negative POCUS findings in high-risk presentations should not preclude definitive imaging with CT, given that 5% false-negative rate (95% sensitivity) translates to missed diagnoses in around one in 20 true AAA cases. Second, diameter quantification using POCUS showed systematic bias requiring either standardized outer-to-outer measurement conventions or recognition that POCUS-derived diameters may not accurately reflect CT measurements. Third, the evidence base derived entirely from settings with at least minimal operator training, and performance in completely untrained operators remains unknown. Fourth, the perfect rupture detection rate suggests high reliability for this important outcome; however, physicians should maintain awareness that small sample size limits confidence in this estimate.

For HUD applications, the evidence remains insufficient to formulate definitive recommendations nor conclusions yet. The single included study showed promising accuracy; however, wide 95% CI, small sample size, and limited generalizability preclude strong conclusions. Primary care physicians or community settings considering HUD implementation for AAA screening should recognize that current evidence quality remains LOW, and decisions should incorporate local expertise, training availability, and alternative screening options. The excellent measurement agreement observed for HUD suggests possible advantages over cart-based systems for diameter quantification; however, this finding requires validation in larger samples across different settings and operators.

Future studies are highly recommended to prioritize several key areas. First, large-scale multicenter trials enrolling symptomatic ED patients are needed to narrow CIs for sensitivity and specificity estimates, with target sample sizes of at least 200 disease-positive patients to achieve precise (±5%) estimation. Second, dedicated studies investigating HUD diagnostic accuracy in different settings including primary care screening, community-based detection, and resource-limited environments would address important evidence gaps. Third, standardized training protocols with clearly defined minimum requirements should be developed and validated across different experience levels. Fourth, measurement agreement studies should universally adopt outer-to-outer conventions aligned with reference standards to enable pooling and eliminate systematic bias. Fifth, cost-effectiveness analyses comparing POCUS-based pathways to standard CT-based approaches would inform implementation decisions and resource allocation.

5. CONCLUSIONS

This systematic review and meta-analysis have demonstrated that ED POCUS achieved excellent diagnostic accuracy for AAA detection, with pooled sensitivity of 95.0% and specificity of 97.7% across four studies enrolling 533 patients, accompanied by minimal heterogeneity ($I^2 = 0$-11%). Perfect rupture detection was observed across 21 cases (100% sensitivity, 95% CI: 84.5-100%), supporting utility for time-sensitive scenarios. However, measurement agreement analysis revealed significant heterogeneity ($I^2 = 98\%$) driven by inner-to-inner versus outer-to-outer measurement convention discordance, necessitating standardized approaches for diameter quantification. Evidence for HUD remained limited to single-study data (104 patients) with wide 95% CI and LOW GRADE certainty, precluding definitive conclusions. Important limitations included small sample sizes, possible publication bias from limited studies, and evidence restricted to symptomatic ED populations. While findings support possible benefits of POCUS for ED-based AAA detection, physicians should maintain awareness of limitations including possible false-negative results, measurement bias, and training requirements. Cautious implementation with recognition of evidence gaps represents a reasonable approach pending additional large-scale validation studies.

REFERENCES

[1] Howard DPJ, Banerjee A, Fairhead JF, et al, Study TOV. Population-Based Study of Incidence of Acute Abdominal Aortic Aneurysms With Projected Impact of Screening Strategy. *Journal of the American Heart Association*. Published online August 25, 2015. DOI:10.1161/jaha.115.001926

[2] Shaban EE, Yigit Y, Alkahlout B, et al, Zaki HA. Enhancing clinical outcomes: Point of care ultrasound in the precision diagnosis and Management of Abdominal Aortic Aneurysms in emergency medicine: A systematic review and meta-analysis. *Journal of Clinical Ultrasound*. Published online September 29, 2024. DOI:10.1002/jcu.23850

[3] Chaikof EL, Dalman RL, Eskandari MK, et al, Starnes BW. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *Journal of Vascular Surgery*. Published online January, 2018. DOI:10.1016/j.jvs.2017.10.044

[4] Wanhainen A, Van Herzeele I, Bastos Goncalves F, et al, Yeung KK. Editor's Choice -- European Society for Vascular Surgery (ESVS) 2024 Clinical Practice Guidelines on the Management of Abdominal Aorto-Iliac Artery Aneurysms. *European Journal of Vascular and Endovascular Surgery*. Published online February, 2024. DOI:10.1016/j.ejvs.2023.11.002

[5] Choi W, Cho YS, Ha YR, et al, on behalf of the Society Emergency and Critical Care Imaging (SECCI). Role of point-of-care ultrasound in critical care and emergency medicine: update and future perspective. *Clinical and Experimental Emergency Medicine*. Published online December 29, 2023. DOI:10.15441/ceem.23.101

[6] Starck J, Brunkwall S, Gottsäter A, et al, Holst J. Evaluation Of Handheld Point-of-Care-Ultrasound (POCUS) in Screening for Abdominal Aortic Aneurysm in Men: A Comparative, Cohort Study. *Ultrasound in Medicine & Biology*. Published online October, 2025. DOI:10.1016/j.ultrasmedbio.2025.07.004

[7] Gibbons RC, Jaeger D, Berger M, et al, Costantino T. Diagnostic Accuracy of a Handheld Ultrasound vs a Cart-based Model: A Randomized Clinical Trial. *Pubmed*. Published online March 01, 2024. DOI:10.5811/westjem.17822

[8] Rykkje A, Carlsen JF, Nielsen MB. Hand-Held Ultrasound Devices Compared with High-End Ultrasound Systems: A Systematic Review. *Diagnostics*. Published online June 15, 2019. DOI:10.3390/diagnostics9020061

[9] Rubano E, Mehta N, Caputo W, et al, Sinert R. Systematic Review: Emergency Department Bedside Ultrasonography for Diagnosing Suspected Abdominal Aortic Aneurysm. *Academic Emergency Medicine*. Published online February 13, 2013. DOI:10.1111/acem.12080

[10] Page MJ, McKenzie JE, Bossuyt PM, et al, Moher D. Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española De Cardiología (english Edition)*. Published online September, 2021. DOI:10.1016/j.rec.2021.07.010

[11] Weber F, Knapp G, Ickstadt K, et al, Ä G. Zero-cell corrections in random-effects meta-analyses. *Research Synthesis Methods*. Published online October 21, 2020. DOI:10.1002/jrsm.1460

[12] Wei JJ, Lin EX, Shi JD, et al, Tong TJ. Meta-analysis with zero-event studies: a comparative study with application to COVID-19 data. *Military Medical Research*. Published online July 03, 2021. DOI:10.1186/s40779-021-00331-6

[13] Bloom B, Gibbons R, Brandis D, et al, Costantino T. Point-of-care Ultrasound Diagnosis of Acute Abdominal Aortic Occlusion. *Clinical Practice and Cases in Emergency Medicine*. Published online January 23, 2020. DOI:10.5811/cpcem.2019.11.44311

[14] Seu R, Gartenberg A, Mirsky R, et al, Halperin M. Point of Care Ultrasound (POCUS) Used to Rapidly Diagnose Both Renal Colic and a Symptomatic Abdominal Aortic Aneurysm in an Elderly Man with Left Flank Pain. *Pocus Journal*. Published online April 15, 2025. DOI:10.24908/pocusj.v10i01.18461

[15] Fatima M, Farooq M, Butt MSJ, et al, Irshad A. Evaluating the Diagnostic Accuracy of Point-of-Care Ultrasound in Acute Abdominal Pain in the Emergency Setting. *Cureus*. Published online September 22, 2025. DOI:10.7759/cureus.92916

[16] Gupta A, Kindarara DM, Chun KC, et al, Lee ES. Accuracy of Point-of-Care Ultrasound in Follow Up Abdominal Aortic Aneurysm Imaging. *Vascular and Endovascular Surgery*. Published online April 28, 2022. DOI:10.1177/15385744221099093

[17] Hartrich M, Eilbert W. An Approach to Point-Of-Care Ultrasound Evaluation of the Abdominal Aorta. *Journal of Visualized Experiments*. Published online September 08, 2023. DOI:10.3791/65487

[18] Radonjić T, Popović M, Zdravković M, et al, Branković M. Point-of-Care Abdominal Ultrasonography (POCUS) on the Way to the Right and Rapid Diagnosis. *Diagnostics*. Published online August 24, 2022. DOI:10.3390/diagnostics12092052

[19] Ienghong K, Khemtong S, Cheung LW, et al, Apiratwarakul K. Enhanced Diagnostic Precision in Revisited Emergency Patients via Point-of-Care Ultrasound: A Tool for Emergency Department Quality Management. *Journal of Multidisciplinary Healthcare*. Published online March, 2025. DOI:10.2147/jmdh.s507075

[20] Tsujimoto Y, Tsutsumi Y, Kataoka Y, et al, Furukawa TA. The impact of continuity correction methods in Cochrane reviews with single-zero trials with rare events: A meta-epidemiological study. *Research Synthesis Methods*. Published online May 15, 2024. DOI:10.1002/jrsm.1720

[21] Zabriskie BN, Cole N, Baldauf J, et al, Decker C. The impact of correction methods on rare-event meta-analysis. *Research Synthesis Methods*. Published online November 09, 2023. DOI:10.1002/jrsm.1677

[22] M JS, A JS, P CL. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*. Published online April 06, 2004. DOI:10.1002/sim.1761

[23] Alfuraih AM, Alrashed AI, Almazyad SO, et al, Alsaadi MJ. Abdominal aorta measurements by a handheld ultrasound device compared with a conventional cart-based ultrasound machine. *Annals of Saudi Medicine*. Published online November, 2021. DOI:10.5144/0256-4947.2021.376

[24] Sisó-Almirall A, Kostov B, Navarro González M, et al, Benavent Àreu J. Abdominal aortic aneurysm screening program using hand-held ultrasound in primary healthcare. *Plos One*. Published online April 28, 2017. DOI:10.1371/journal.pone.0176877

[25] Bonnafy T, Lacroix P, Desormais I, et al, Aboyans V. Reliability of the measurement of the abdominal aortic diameter by novice operators using a pocket-sized ultrasound system. *Archives of Cardiovascular Diseases*. Published online December, 2013. DOI:10.1016/j.acvd.2013.08.004

[26] Moore CL, Holliday RS, Hwang JQ, et al, Osborne MR. Screening for abdominal aortic aneurysm in asymptomatic at-risk patients using emergency ultrasound. *The American Journal of Emergency Medicine*. Published online October, 2008. DOI:10.1016/j.ajem.2007.11.030

[27] Dent B, Kendall RJ, Boyle AA, et al, Atkinson PRT. Emergency ultrasound of the abdominal aorta by UK emergency physicians: a prospective cohort study. *Emergency Medicine Journal*. Published online August 01, 2007. DOI:10.1136/emj.2007.048405

[28] Knaut AL, Kendall JL, Patten R, et al, Ray C. Ultrasonographic measurement of aortic diameter by emergency physicians approximates results obtained by computed tomography. *The Journal of Emergency Medicine*. Published online February, 2005. DOI:10.1016/j.jemermed.2004.07.013

[29] Costantino TG, Bruno EC, Handly N, et al, Dean AJ. Accuracy of emergency medicine ultrasound in the evaluation of abdominal aortic aneurysm. *The Journal of Emergency Medicine*. Published online November, 2005. DOI:10.1016/j.jemermed.2005.02.016

[30] Tayal VS, Graf CD, Gibbs MA. Prospective Study of Accuracy and Outcome of Emergency Ultrasound for Abdominal Aortic Aneurysm over Two Years. *Academic Emergency Medicine*. Published online June 28, 2008. DOI:10.1197/aemj.10.8.867

[31] Lin PH, Bush RL, McCoy SA, et al, Lumsden AB. A prospective study of a hand-held ultrasound device in abdominal aortic aneurysm evaluation. *The American Journal of Surgery*. Published online November, 2003. DOI:10.1016/j.amjsurg.2003.07.018

[32] Kuhn M, Bonnin RLL, Davey MJ, et al, Langlois SLP. Emergency Department Ultrasound Scanning for Abdominal Aortic Aneurysm: Accessible, Accurate, and Advantageous. *Annals of Emergency Medicine*. Published online September, 2000. DOI:10.1067/mem.2000.108616

APPENDICES

Supplementary Table 1: Risk of Bias and Applicability Assessment of Included Studies.

Study	Patient Selection		Index Test		Reference Standard		Flow and Timing	Overall Risk
	Risk of Bias	Applicability	Risk of Bias	Applicability	Risk of Bias	Applicability	Risk of Bias	
Alfuraih et al. 2021	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	Low risk
Sisó-Almirall et al. 2017	⊖ Low	⊕ High	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊕ High	High risk
Bonnafy et al. 2013	⊖ Low	⊖ Low	⊖ Low	? Unclear	⊖ Low	⊖ Low	⊖ Low	Low risk
Moore et al. 2008	⊖ Low	⊕ High	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊕ High	High risk
Dent et al. 2007	? Unclear	⊖ Low	⊖ Low	⊖ Low	? Unclear	⊖ Low	⊕ High	High risk
Knaut et al. 2005	⊖ Low	⊖ Low	⊕ High	⊖ Low	⊖ Low	⊖ Low	⊖ Low	Moderate risk
Costantino et al. 2005	⊖ Low	⊖ Low	? Unclear	⊖ Low	⊖ Low	⊖ Low	⊖ Low	Low risk
Tayal et al. 2003	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	Low risk
Lin et al. 2003	⊖ Low	? Unclear	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊖ Low	Low risk
Kuhn et al. 2000	⊖ Low	⊖ Low	⊖ Low	⊖ Low	⊕ High	⊖ Low	? Unclear	High risk

Abbreviations: ITI, inner-to-inner; OTO, outer-to-outer; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.

Supplementary Table 2: Summary of Findings and GRADE Certainty Assessment.

Outcome	Studies (Patients)	Effect Estimate (95% CI)	Absolute Effect per 1000 Patients[a]	Certainty	Factors Affecting Certainty
Emergency Department POCUS (Symptomatic, Prevalence 15%):					
Sensitivity for AAA ≥3.0 cm	4 (533)	95.0% (87.5-98.1)	143 detected / 150 with AAA	⊕⊕⊕⊕ HIGH[b]	No downgrading; continuity correction provides realistic uncertainty
Specificity for AAA ≥3.0 cm	4 (533)	97.7% (94.9-99.0)	830 correctly ruled out / 850 without AAA	⊕⊕⊕⊕ HIGH[b]	No downgrading; low heterogeneity (I^2=11%), robust across studies
Mean diameter difference	4 (251)	Unpoolable (range: -2.6 to +4.4 mm)	-	⊕⊕○○ LOW[c]	Downgraded twice: serious inconsistency (I^2=98%, ITI vs OTO bias), imprecision
Rupture detection sensitivity	3 (21)	100.0% (84.5-100.0)	21/21 detected	⊕⊕⊕○ MODERATE[d]	Downgraded once: serious imprecision (small sample, N=21 events)
Handheld Ultrasound (Vascular Clinic/Primary Care):					
Diagnostic accuracy (≥3.0 cm)	1 (104)	Se 90.6% (67.7-97.8), Sp 96.1% (89.9-98.6)	14/15 detected	⊕⊕○○ LOW[e]	Downgraded twice: very serious imprecision (single study, wide CI), indirectness
Mean diameter difference	3 (284)	0.51 mm (0.29-0.72)	-	⊕⊕⊕⊕ HIGH[f]	No downgrading; low heterogeneity (I^2=11%), tight CI, clinically negligible bias
ICC (reliability)	3 (284)	0.865 (0.832-0.892)	-	⊕⊕⊕⊕ HIGH[f]	No downgrading; excellent reliability, narrow CI, consistent across studies
Within ±3 mm proportion	3 (284)	91.4%	-	⊕⊕⊕○ MODERATE[g]	Downgraded once: imprecision (limited data, only 3 measurements)
Secondary Outcomes:					
Screening yield (HUD)	1 (1010)	2.49% detection rate	25/1000 screened	⊕○○○ VERY LOW[h]	Downgraded three times: risk of bias (partial verification), imprecision, indirectness
Incomplete visualization (ED)	4 (533)	0-37.5% (median 2.9%)	-	⊕⊕○○ LOW[i]	Downgraded twice: inconsistency (high variability), reporting bias
Scan time (ED/HUD)	3-6 (174-515)	2.4-5.3 min (range)	-	⊕⊕○○ LOW[i]	Downgraded twice: inconsistency, incomplete reporting

Abbreviations: AAA, abdominal aortic aneurysm; CI, confidence interval; ED, emergency department; GRADE, Grading of Recommendations Assessment, Development and Evaluation; HUD, handheld ultrasound device; ICC, intraclass correlation coefficient; ITI, inner-to-inner; OTO, outer-to-outer; POCUS, point-of-care ultrasound; Se, sensitivity; Sp, specificity. Notes: GRADE Certainty: ⊕⊕⊕⊕ HIGH | ⊕⊕⊕○ MODERATE | ⊕⊕○○ LOW | ⊕○○○ VERY LOW, [a] Based on 15% AAA prevalence (weighted mean from ED studies); [b] HIGH certainty upgraded from previous MODERATE: Continuity correction eliminated numerical artifacts (previous 100% estimates were methodological flaws), realistic CIs now provided, low heterogeneity (I^2=0-11%) validates pooling, robust sensitivity analyses (Table 4), no serious limitations; [c] LOW certainty: Extreme heterogeneity (I^2=98%) driven by ITI vs OTO measurement conventions; unpoolable across mixed conventions; OTO-only subanalysis shows I^2=0%; [d] MODERATE certainty: Perfect detection (21/21) but small sample creates wide CI lower bound (84.5%); no downgrade for consistency or bias; [e] LOW certainty: Single study (Lin 2003), wide CI especially for sensitivity (68-98%), vascular clinic population differs from primary care screening; [f] HIGH certainty upgraded: HUD agreement shows minimal bias (0.5 mm), low heterogeneity (I^2=11%), >90% within ±3 mm, excellent ICC (0.87); robust evidence for clinical use; [g] MODERATE certainty: Limited to 3 measurements from 2 studies, but consistent findings (88-95% range); [h] VERY LOW certainty: Sisó-Almirall partial verification bias (only 14/1010 verified), single study, Spanish males ≥60 may not generalize; [i] LOW certainty: Inconsistent/selective reporting across studies.

ADDITIONAL INFORMATION

Author Contributions

Yazan J Alalwani: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft. Lamees A Alzahrani: Investigation, Validation, Project Administration, Writing - Review & Editing. Maysam H Alyami: Validation, Formal Analysis, Writing - Review & Editing. Abdulrahman A Alanazi: Validation, Formal Analysis, Writing - Review & Editing. Mawaddah A Alsobay: Data Curation, Investigation, Writing - Review & Editing. Ibtihaj A Alsalhi: Methodology, Writing - Original Draft, Software, Visualization. Alhawraa H Alawamia: Methodology, Software, Visualization, Writing - Review & Editing. Jumanah S Al-Siari: Writing - Review & Editing, Resources, Data Curation. Reema A Alharbi: Formal Analysis, Writing - Review & Editing. Dalya M Motabagani: Formal Analysis, Writing - Review & Editing. Ahmed Y Azzam: Conceptualization, Writing - Review & Editing, Supervision, Project Administration

Human Ethics

IRB Status: not_applicable Consent: Not applicable

Animal Ethics

This study did not involve animal subjects or tissue.

Conflicts of Interest

No conflicts of interest to disclose.

Funding

No funding was received for this study.

AI Use Disclosure

During the preparation of this manuscript, the author(s) used AI-assisted tools for language editing purposes, including grammar correction, clarity refinement, and readability improvement. After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the accuracy and integrity of all content in the final manuscript. No AI tools were used to generate, replace, or alter scientific content, data, analysis, or conclusions.

Acknowledgements

None.

HOW TO CITE THIS ARTICLE

Alalwani YJ, Alzahrani LA, Alyami MH, et al, Azzam AY. **Diagnostic Accuracy of Point-of-Care and Handheld Ultrasound for Abdominal Aortic Aneurysm Detection: A Systematic Review and Meta-Analysis**. *Evidance Health Sciences*. Published online May 31, 2026. DOI:**10.65416/ehealthsci.2026.673710**